Chapter 6

Quasi Monte Carlo methods

In this chapter we discuss the use of low-discrepancy sampling to replace the pure random sampling that forms the backbone of the Monte Carlo method. Using this alternative sampling method in the context of multivariate integration is usually referred to as *quasi-Monte Carlo*. A low-discrepancy sample is one whose points are distributed in a way that approximates the uniform distribution as closely as possible. Unlike for random sampling, points are not required to be independent. In fact, the sample might be completely deterministic. There are several techniques to build *good* point sets; here we only provide a short introduction, since these techniques stand at the basis of relevant *sensitivity analysis* and *uncertainty quantification* methods.

we cannot use CLT (for instance)

6.1 Idea and Construction

We consider the same setting as for LHS, namely the goal of computing $\mu = \mathbb{E}[Z]$, with $Z = \Psi(X_1, \dots, X_d)$ and $\mathbf{X} = (X_1, \dots, X_d) \sim \mathcal{U}([0,1]^d)$, hence computing μ turns into a high-dimensional integration problem,

$$\mu = \int_{[0,1]^d} \Psi(x_1, \dots, x_d) dx_1 \dots dx_d.$$

A crude Monte Carlo estimators using N replicas achieves an error

$$|\mu - \hat{\mu}_{CMC}| \leq z_{1-\frac{\alpha}{2}} \frac{\sqrt{\mathrm{Var}[Z]}}{\sqrt{N}}$$

with asymptotic confidence $1 - \alpha$.

The idea of Quasi Monte Carlo sampling is to turn to a *pure deterministic* sampling to improve the rate $1/\sqrt{N}$, while keeping the simple structure of the sample average estimator,

$$\hat{\mu}_{QMC} = \frac{1}{N} \sum_{i=1}^{N} \Psi(\mathbf{x}^{(j)})$$

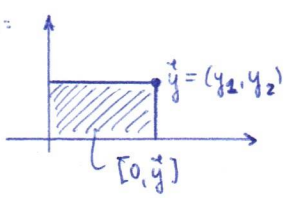
all the focus on: how can we generate good points sets $(\vec{x}^{(j)})$? ("good" in terms of discrepancy)

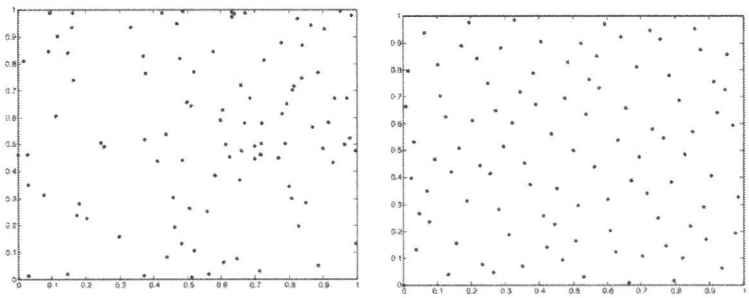
with equal weights $1/N$. Quasi Monte Carlo sampling relies on the observation that a random sampling does not look so uniform in the unit hypercube and hopefully there exist better designs that cover more uniformly the unit cube. Hence, if the Monte Carlo method chooses randomly (or pseudorandomly) distributed points, the quasi-Monte Carlo method chooses points as elements of a low-discrepancy sequence, in the sense precised below.

The main notion in this respect is that of *discrepancy*. We introduce the following notation: for a point $\mathbf{y} \in [0,1]^d$, $\mathbf{y} = (y_1, \dots, y_d)$, we denote by $[0, \mathbf{y}]$ the hyper-rectangle

$$[0, \mathbf{y}] = \prod_{i=1}^{d} [0, y_i]$$

($d=2$):





(a) (b)

Figure 6.1: Left: random sampling; right: QMC sampling (Sobol sequence)

with volume $Vol([0,\mathbf{y}]) = \prod_{i=1}^{d} y_i$.

For an arbitrary sample of N points in $[0,1]^d$, $\mathcal{P} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$ hereafter called a *point set*, we introduce the volume estimator

$$\hat{Vol}_{\mathcal{P}}([0,\mathbf{y}]) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{[0,\mathbf{y}]}(\mathbf{x}^{(i)}) = \frac{\#\{\mathbf{x}^{(i)} \in [0,\mathbf{y}]\}}{N};$$

we introduce an approximation of $Vol([0,\vec{y}])$ and then we compare it with $Vol([0,\vec{y}])$. The comparison (i.e. the difference) is the DISCREPANCY

$\hat{Vol}_{\mathcal{P}}([0,\mathbf{y}])$ is an empirical estimation of $Vol([0,\mathbf{y}])$ based on the point set $\mathcal{P}$.

Definition 6.1.1. *We define* discrepancy function $\Delta_{\mathcal{P}} : [0,1]^d \to [0,1]$ *the function*

$$\Delta_{\mathcal{P}}(\mathbf{y}) = \hat{Vol}_{\mathcal{P}}([0,\mathbf{y}]) - Vol([0,\mathbf{y}]) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{[0,\mathbf{y}]}(\mathbf{x}^{(i)}) - \prod_{i=1}^{d} y_i.$$

From the discrepancy function, several measures of discrepancy of a point set P can be derived; here we mention the two most common examples, namely:

measures of how well the points fit the hyper-rectangle

- L_q-discrepancy,

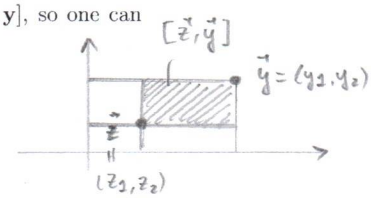
$$D_{N,q}(\mathcal{P}) = \|\Delta_{\mathcal{P}}\|_{L^q} = \left(\int_{[0,1]^d} |\Delta_{\mathcal{P}}(\mathbf{y})|^q d\mathbf{y}\right)^{1/q};$$

- star-discrepancy,

$$D_N^*(\mathcal{P}) = \|\Delta_{\mathcal{P}}\|_{L^\infty} = \sup_{\mathbf{y}\in[0,1]^d} |\Delta_{\mathcal{P}}(\mathbf{y})|.$$

There is actually nothing special in choosing only the rectangles of the form $[0,\mathbf{y}]$, so one can define also the extreme discrepancy:

$$D_N(\mathcal{P}) = \sup_{\mathbf{y},\mathbf{z}\in[0,1]^d \,:\mathbf{z}\le\mathbf{y}} |\hat{Vol}_{\mathcal{P}}([\mathbf{z},\mathbf{y}]) - Vol([\mathbf{z},\mathbf{y}])|.$$



It can be (easily) shown that

$$D_N^*(\mathcal{P}) \le D_N(\mathcal{P}) \le 2^d D_N^*(\mathcal{P});$$

the left inequality is obvious; the right inequality follows from the observation that a rectangle $[\mathbf{z},\mathbf{y}]$ can be written as composition (union/intersection) of 2^d rectangles of the type $[0,\mathbf{t}]$. Hence,

it suffices to study only the star-discrepancy. With these definitions, discrepancy represents the worst-case or maximum point density deviation of a uniform set.

The reason why the discrepancy plays an important role in the study of QMC quadrature formulas follows from the *Koksma-Hlawka inequality*, which allows us to bound the error of the Quasi Monte Carlo method by the product of two terms, one of which depends only on the function being integrated, and the other one is the discrepancy of the point set $\mathcal{P}$.

Let us first illustrate this inequality in dimension $d = 1$. To this aim, we first state the

(Zaremba's identity). Let $\Psi : [0,1] \to \mathbb{R}$ an absolutely continuous function with integrable derivative, and $\mathcal{P} = \{X^{(1)}, \dots, X^{(N)}\}$ any point set in $[0,1]$. Then

$$\int_0^1 \Psi(x)dx - \frac{1}{N}\sum_{i=1}^{N} \Psi(x^{(i)}) = \int_0^1 \Psi'(y)\Delta_{\mathcal{P}}(y)dy.$$

Proof. We use the identity

$$\Psi(x) = \Psi(1) - \int_x^1 \Psi'(y)dy.$$

Then

$$\int_0^1 \Psi(x)dx - \frac{1}{N}\sum_i \Psi(x^{(i)}) = \Psi(1) - \int_0^1 \int_x^1 \Psi'(y)dydx - \frac{1}{N}\sum_i \Psi(1) + \frac{1}{N}\sum_i \int_{x_i}^1 \Psi'(y)dy.$$

Since, reversing the order of integration (do you remember? :-))

$$\int_0^1 \int_x^1 \Psi'(y)dydx = \int_0^1 \int_0^y \Psi'(y)dxdy = \int_0^1 y\Psi'(y)dy$$

and

$$\int_{x_i}^1 \Psi'(y)dy = \int_0^1 \Psi'(y)\mathbf{1}_{[x_i,1]}(y)dy$$

we end up with

$$\begin{aligned}\int_0^1 \Psi(x)dx - \frac{1}{N}\sum_i \Psi(x^{(i)}) = & -\int_0^1 \int_0^y \Psi'(y)dxdy + \frac{1}{N}\sum_i \int_0^1 \Psi'(y)\mathbf{1}_{[x_i,1]}(y)dy \\ = & \int_0^1 \Psi'(y)\left[\frac{1}{N}\sum_i \mathbf{1}_{[x_i,1]}(y) - y\right]dy \\ = & \int_0^1 \Psi'(y)\left[\frac{1}{N}\sum_i \mathbf{1}_{[0,y]}(x_i) - y\right]dy = \int_0^1 \Psi'(y)\Delta_{\mathcal{P}}(y)dy.\end{aligned}$$

□

From the Zaremba's identity, the following inequality immediately follows, applying the Holder inequality. Recall that

$$W^{1,p}(\mathbb{R}) = \{f \in L^p(\mathbb{R}) \ : \ f' \in L^p(\mathbb{R})\}$$

(Koksma-Hlawka inequality). For any $\Psi : \mathbb{R} \to \mathbb{R}$ such that $\Psi \in W^{1,p}(\mathbb{R})$,

$$\left|\int_0^1 \Psi(x)dx - \frac{1}{N}\sum_{i=1}^{N} \Psi(x^{(i)})\right| \le \|\Psi'\|_{L^p}\|\Delta_{\mathcal{P}}\|_{L^q}.$$

$\left(\frac{1}{p} + \frac{1}{q} = 1\right)$

Hence, we remark that the quadrature error behaves as $\|\Delta_{\mathcal{P}}\|_{L^q}$ provided that $\Psi' \in L^p(\mathbb{R})$, that is, $\Psi \in W^{1,p}(\mathbb{R})$. In particular, if Ψ' is integrable then

$$\left| \int_0^1 \Psi(x)dx - \frac{1}{N}\sum_{i=1}^{N} \Psi(x^{(i)}) \right| \leq \|\Psi'\|_{L^1} D_N^*(\mathcal{P}).$$

The previous analysis extends with some care to the multi-dimensional setting; from the so-called Hlawka's identity – the multi-dimensional version of Zaremba's identity – it is possible to derive the multi-dimensional version of the Koksma-Hlawka inequality. Denote by $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_k\}$ a subset of dimensions, with $\alpha_j \in \{1, \dots, d\}$ without repetition and set $|\boldsymbol{\alpha}| = k$. For $\mathbf{x} \in [0,1]^d$, we denote by $\mathbf{x}_{\boldsymbol{\alpha}} = (x_{\alpha_1}, \dots, x_{\alpha_k}) \in [0,1]^k$ and $\mathbf{z} = (\mathbf{x}_{\boldsymbol{\alpha}}, 1)$ the vector $z_j = x_j$ if $j \in \boldsymbol{\alpha}$, $z_j = 1$ if $j \notin \boldsymbol{\alpha}$. With this notation at hand, we can define, for any function $\Psi : [0,1]^d \to \mathbb{R}$ smooth,

$$\|\Psi\|_{p,p'} = \left(\sum_{\boldsymbol{\alpha} \subset \{1,\dots,d\}} \left(\int_{[0,1]^d} \left| \frac{\partial^{|\boldsymbol{\alpha}|}\Psi}{\partial \mathbf{x}_{\boldsymbol{\alpha}}}(\mathbf{y}_{\boldsymbol{\alpha}}, 1) \right|^p d\mathbf{y}_{\boldsymbol{\alpha}} \right)^{p'/p} \right)^{1/p'}$$

basically we require that Ψ has mixed first derivatives that are integrable

(Koksma-Hlawka inequality). For any $\Psi : \mathbb{R}^d \to \mathbb{R}$ such that $\|\Psi\|_{p,p'} < +\infty$,

$$\left| \int_{[0,1]^d} \Psi(\mathbf{x})d\mathbf{x} - \frac{1}{N}\sum_{i=1}^{N} \Psi(\mathbf{x}^{(i)}) \right| \leq \|\Psi\|_{p,p'} \|\Delta_{\mathcal{P}}\|_{q,q'}$$

with $1/p + 1/q = 1/p' + 1/q' = 1$. In particular, if $\|\Psi\|_{1,1} < +\infty$, then

$$\left| \int_{[0,1]^d} \Psi(\mathbf{x})d\mathbf{x} - \frac{1}{N}\sum_{i=1}^{N} \Psi(\mathbf{x}^{(i)}) \right| \leq \|\Psi\|_{1,1} D_N^*(\mathcal{P}).$$

In practice: the quadrature error behaves as $D_N^*(\mathcal{P})$ provided ψ has integrable mixed first-order derivatives.

Hence, if the integrand is given, and cannot be changed, the inequality motivates the search for quadrature points with small discrepancy. The previous results also connects QMC with the field of discrepancy theory, since it shows that point sets with low discrepancy are attractive to use as QMC integration points.

6.2 Low discrepancy sequences and point sets

There exists constructions of point sets $\mathcal{P} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$, $\mathbf{x}^{(i)} \in [0,1]^d$, that have star-discrepancy as lows as

$$D_N^*(\mathcal{P}) = O\left(\frac{(\log N)^{d-1}}{N} \right).$$

here we have N and no longer $\sqrt{N}$:)
(we can improve MC, however this time we have dependence on the dimension d)

It is widely believed that this result is sharp, that is, there do not exist points set that achieve a better bound. In general, these constructions do not lead to nested sequences of points, that is, the point set $\mathcal{P}' = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}\}$ with $M > N$ in general does not contain $\mathcal{P}$.

For the nested case, i.e. point sets $\mathcal{P} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$ that are generated as the first N points of an infinite sequence $\mathcal{S} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots\}$ the discrepancy results are slightly worse, namely

$$D_N^*(\mathcal{P}) = O\left(\frac{(\log N)^d}{N} \right).$$

Therefore, in the following we refer to a *low discrepancy point set* $\mathcal{P} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$ as one that achieves the bound $D_N^*(\mathcal{P}) = O\left(\frac{(\log N)^{d-1}}{N}\right)$, and to a *low discrepancy sequence* $\mathcal{S} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots\}$ as one for which the corresponding point sets $\mathcal{P}$ achieve the bound $D_N^*(\mathcal{P}) = O\left(\frac{(\log N)^d}{N}\right)$.

From these considerations, we see that a QMC quadrature formula can achieve rate $1/N$ up to logarithmic terms (which however grow exponentially in the dimension d!), provided the integrand function has integrable mixed first derivatives.

Take-home Message

Quasi-Monte Carlo (QMC) methods take the same form as the Monte Carlo method in the unit cube, but instead of generating the sample points (pseudo)randomly, we choose them deterministically in a clever way to be more uniformly distributed than random points, so that they have a faster rate of convergence.

All QMC theoretical error bounds take the common form of a product

$$\left| \int_{[0,1]^d} \Psi(\mathbf{x}) d\mathbf{x} - \frac{1}{N} \sum_{i=1}^{N} \Psi(\mathbf{x}^{(i)}) \right| \leq D_N^*(\mathcal{P}) V(\Psi),$$

with one factor depending only on the points and the other depending only on the integrand. In the classical theory these two factors are called the *discrepancy* of the points set and the *variation* of Ψ, respectively.

If the integrand Ψ has sufficient smoothness, e.g., can be differentiated once with respect to each variable, then classical theory tells us that certain QMC methods can converge like $O((\log N)^d/N)$; they are referred to as low-discrepancy sequences.

The drawback of the classical QMC theory is that the error bound and implied constant grow exponentially with dimension d, so the theory is not useful when d is very large.

Remark 6.2.1. *Because any distribution of random numbers can be mapped onto a uniform distribution, and quasi-random numbers are mapped in the same way, we are only concerned with the generation of quasi-random numbers on a multidimensional uniform distribution. The same holds for integration: by means of a change of variable, we are always able to transform an integral over* $\mathbb{R}^d$ *to an integral over* $[0,1]^d$.

Before presenting some examples, let us give two important remarks:

1. Consider the point set $\mathcal{P} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$ with $\mathbf{x}^{(i)} \overset{i.i.d.}{\sim} \mathcal{U}([0,1]^d)$, that is a *random* i.i.d. sample. In this case

$$|\Delta_{\mathcal{P}}(\mathbf{y})| = |\hat{Vol}_{\mathcal{P}}([0,\mathbf{y}]) - Vol([0,\mathbf{y}])| \approx O\left(\frac{1}{\sqrt{N}}\right)$$

   in a probabilistic sense – indeed, $|\Delta_{\mathcal{P}}(\mathbf{y})|$ is the error of the sample average estimator of $Vol_{\mathcal{P}}([0,\mathbf{y}])$; a random i.i.d. sequence is not a low-discrepancy sequence.

2. Consider a regular lattice

$$\mathcal{P} = \left\{ \left(\frac{k_1 + 1/2}{m}, \dots, \frac{k_d + 1/2}{m} \right), \ 0 \leq k_j \leq m-1, \ j = 1, \dots, d \right\}, \quad N = m^d.$$

If $d = 1$, then $D_N^*(\mathcal{P}) = 1/2m$. In arbitrary dimension $d > 1$

$$D_N^*(\mathcal{P}) = \sup_{\mathbf{y}\in[0,1]^d} |\Delta_{\mathcal{P}}(\mathbf{y})| \geq \sup_{t\in[0,1]} |\Delta_{\mathcal{P}}(t, 1, \dots, 1)| = \frac{1}{2m} = \frac{1}{2N^{1/d}}.$$

Hence, again, a regular lattice is not a low-discrepancy point set.

Let us now consider some common low-discrepancy sequences/points sets.

some examples (we don't go into det.)

Van der Corput sequence
Let $b \geq 2$ an integer. Any natural number $n \in \mathbb{N}_0$ can be expanded in b-adic expansion

$$n = n_1 + n_2 b + n_3 b^2 + \dots = \sum_{i=1}^{\infty} n_i b^{i-1}, \qquad n_i \in \{0, 1, \dots, b-1\}$$

the *radical inverse* of n is defined as

$$\varphi_b(n) = \frac{n_1}{b} + \frac{n_2}{b^2} + \dots = \sum_{i=1}^{\infty} \frac{n_i}{b^i}.$$

In other words, if $n = (\dots n_2\, n_1)_b$ denotes the base b representation of n, then $\varphi_b(n) = (0.n_1 n_2 \dots)_b$.

In dimension $d = 1$, the *van der Corput sequence* in base b is the one-dimensional sequence

$$\varphi_b(0), \varphi_b(1), \varphi_b(2), \dots$$

For example, take $b = 2$. First we write down the natural numbers $0, 1, 2, \dots$ in base 2:

$$0, 1_2, 10_2, 11_2, 100_2, 101_2, 110_2, \dots$$

Then we apply the radical inverse function $\varphi_2(\cdot)$ to each number, to obtain the sequence

$$0, 0.1_2, 0.01_2, 0.11_2, 0.001_2, 0.101_2, 0.011_2, \dots,$$

which in decimal form is the sequence

$$0, 0.5, 0.25, 0.75, 0.125, 0.625, 0.375, \dots.$$

Halton sequence
The *Halton sequence* generalizes this construction for $d \geq 2$. Let $p_1, p_2, \dots, p_s$ be the first s prime numbers. The Halton sequence $\mathcal{S} = \{\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots\}$ in s dimensions is given by

$$\mathbf{x}^{(n)} = (\varphi_{p_1}(n), \varphi_{p_2}(n), \dots, \varphi_{p_s}(n)), \ n = 0, 1, \dots$$

that is, the j-th components of points in the Halton sequence form the van der Corput sequence in base p_j, where p_j is the j-th prime. The Halton sequence leads to an *open* QMC method, and achieve the bound

$$D_N^*(\mathcal{P}) \leq C(d) \frac{(\log N)^d}{N} \qquad \forall d \geq 2.$$

where $C(d)$ only depends on d. Observe that, although the convergence rate appears to beat the MC rate of $O(n^{-1/2})$, the MC rate is independent of d, whereas for fixed d, the function $(\log D)^d/N$ increases with increasing N for all $N < e^d$.

Remark 6.2.2. *As a matter of fact, there are two types of QMC methods.*

- *the* open *type: this uses the first N points of an infinite sequence. Thus to increase N one only needs to evaluate the integrand at the additional quadrature points;*

- *the* closed *type: this uses a finite point set which depends on N. Thus a new value of N means a completely new set of quadrature points.*

Hammersley point set
Let $p_1, p_2, \ldots, p_{s-1}$ be the first $s-1$ prime numbers. The Hammersley point set with N points in s dimensions, $\mathcal{S} = \{\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}\}$, is given by

$$\mathbf{x}^{(n)} = \left(\frac{n}{N}, \varphi_{p_1}(n), \varphi_{p_2}(n), \ldots, \varphi_{p_{s-1}}(n)\right), \quad n = 0, 1, \ldots, N-1.$$

The Hammersley point set leads to a 'closed' QMC method – this Hammersley point sets are not nested ad achieve the better bound

$$D_N^*(\mathcal{P}) \leq C(d)\frac{(\log N)^{d-1}}{N} \qquad \forall d \geq 2.$$

Note that there is one less power of $\log N$ compared to the error bound for the Halton sequence. Typically, the error bounds for QMC methods based on *closed* point sets are better than those based on *open* sequences.

Lattice point set
Let $N \in \mathbb{N}$ and $\mathbf{g} \in \mathbb{N}^d$, $\mathbf{g} = (g_1, \ldots, g_d)$ such that g_i has no factor in common with N (typically N is taken as a prime number), is known as the generating vector. Then, the *lattice point set* (also denoted as N-point rank-one lattice rule[1] in d dimensions) is

$$\mathcal{P} = \left\{\frac{n\mathbf{g}}{N}\right\}_{n=0}^{N-1}$$

where the braces indicate that we take the fractional part of a real number, that is, $\{x\} = x - \lfloor x \rfloor$. Good choices of $\mathbf{g}$ lead to low-discrepancy point sets. Two instances are reported in Figure 6.2.

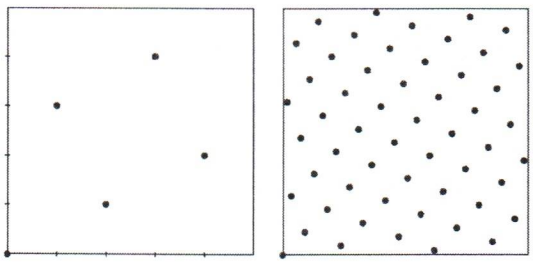


Figure 6.2: A lattice of $N = 5$ points obtained from $\mathbf{g} = (1, 3)$ (left) and a lattice of $N = 55$ points obtained from $\mathbf{g} = (1, 34)$ (right).

(t, m, d)-nets and (t, d) sequences
Nets and sequences provide another method for obtaining well-distributed point sets in the unit cube $[0, 1)$s that are useful for QMC integration. The concept of (t, m, d)-net is based on sub-dividing the unit cube into intervals and placing points in the cube such that each interval of a certain size and shape contains the *correct* number of points.

[1]The minimal number of generating vectors required to generate a lattice rule is known as the rank of the rule.

Let $t \geq 0$, $m \geq 1$, $d \geq 1$, and $b \geq 2$ be integers with $t \leq m$. A (t, m, d)-net in base b is a point set $\mathcal{P}$ consisting of $N = b^m$ points in $[0,1)^d$ such that every elementary rectangle of the form

$$\prod_{j=1}^{d} \left[\frac{a_j}{b^{p_j}}, \frac{a_j + 1}{b^{p_j}} \right)$$

with each $p_j \geq 0$, $0 \leq a_j < b^{p_j}$, and $p_1 + p_2 + \ldots + p_s = m - t$, contains exactly b^t points of $\mathcal{P}$. An elementary rectangle of this form has volume $b^{-(p_1+p_2+\ldots+p_s)} = b^{t-m}$, which is precisely the proportion of the points from $\mathcal{P}$ that lie in this elementary interval. One would expect such a property to hold if the point set is uniformly distributed. Figure 6.3 provides an illustration of a two-dimensional net with 16 points.

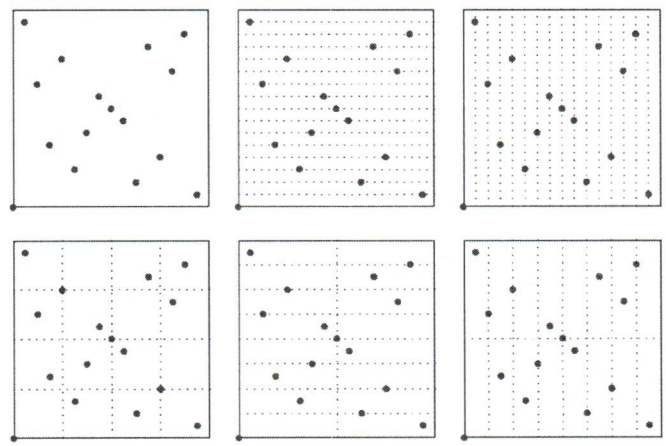


Figure 6.3: Illustration of a $(0, 4, 2)$-net in base 2: this is a point set with $N = 2^4 = 16$ points, every elementary rectangle of volume $2^{-(m-b)} = 2^{-4} = 1/16$ contains exactly $2^t = 2^0 = 1$ point. A point that lies on the dividing line counts toward the interval above or to the right.

There is an analogous concept for infinite sequences. A (t, d)-sequences in base b is a sequence $\mathcal{S} = \{\mathbf{x}_0, \mathbf{x}_1, \ldots\}$ in $[0,1)^d$ such that for any integer $m > t$ and $l \geq 0$, every block of b^m points $\{\mathbf{x}_{lb^m}, \mathbf{x}_{(l+1)b^m-1}\}$ in the sequence $\mathcal{S}$ forms a (t, m, d)-net in base b.

The star-discrepancy of a (t, m, d)-net satisfies

$$D_N^*(\mathcal{P}) = O\left(b^t \frac{(\log N)^{d-1}}{N} \right)$$

and similarly for a (t, d)-sequence we have

$$D_N^*(\mathcal{P}) = O\left(b^t \frac{(\log N)^{d}}{N} \right).$$

Famous (t, d)-sequences are the Sobol' sequence – which is a (t, d)-sequence in base 2 – the Niederreiter sequence, the Faure sequence.

Hands-On 7 focuses on the use of Quasi Monte Carlo formulas for the numerical approximation of integrals.

6.3 Randomized QMC formulas

Let us consider a point set $\mathcal{P} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$ and the QMC quadrature formula

$$\hat{\mu}_{QMC} = \frac{1}{N} \sum_{i=1}^{N} \Psi(\mathbf{x}^{(i)}).$$

The question is how to estimate the error $|\mu - \hat{\mu}_{QMC}|$. Since the points $\mathbf{x}^{(i)}$ are not random i.i.d., we cannot use a variance estimator or a CLT as in the Monte Carlo estimator.

An easy idea to estimate the error is to randomize the QMC formula. Let $\mathbf{U} \sim \mathcal{U}([0,1]^d)$. If $\mathcal{P} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$ is a low-discrepancy point set, so is

$$\mathcal{P}_U\{\{\mathbf{x}^{(1)} + \mathbf{U}\}, \dots, \{\mathbf{x}^{(N)} + \mathbf{U}\}\}$$

where again $\{\cdot\}$ denotes the fractional part. $\mathcal{P}_U$ is called a randomly shifted point set. We could then compute $\hat{\mu}_{QMC}$ for few randomly shifted point sets and average over the obtained results. In other terms, we apply a Monte Carlo estimator on $\hat{\mu}_{QMC}$ with random shift.

$\{x\} := x - \lfloor x \rfloor$
(fractional part)

Algorithm 22 Randomly shifted QMC

1: Generate $\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(k)} \overset{i.i.d.}{\sim} \mathcal{U}([0,1]^d)$
2: **for** $j = 1, \dots, k$ **do**
3: Compute

$$\hat{\mu}_{QMC}^{(j)} = \frac{1}{N} \sum_{i=1}^{N} \Psi(\{\mathbf{x}^{(i)} + \mathbf{U}^{(j)}\})$$

4: **end for**
5: Compute

$$\hat{\mu}_{QMC} = \frac{1}{k} \sum_{j=1}^{k} \hat{\mu}_{QMC}^{(j)} = \frac{1}{Nk} \sum_{j=1}^{k} \sum_{i=1}^{N} \Psi(\{\mathbf{x}^{(i)} + \mathbf{U}^{(j)}\})$$

Since $\mathbf{U}^{(j)} \sim \mathcal{U}([0,1]^d)$, so is $\{\mathbf{x}^{(i)} + \mathbf{U}^{(j)}\}$ for any $i = 1, \dots, N$. It follows that $\hat{\mu}_{QMC}$ is an *unbiased* estimator of $\mu = \mathbb{E}[\Psi]$. Moreover, since $\hat{\mu}_{QMC}^{(j)}$ are independent, the variance of $\hat{\mu}_{QMC}$ is

$$\mathrm{Var}\left[\hat{\mu}_{QMC}\right] = \frac{\sigma_{QMC}^2}{k} \qquad \text{with} \quad \sigma_{QMC}^2 = \mathbb{E}\left[(\hat{\mu}_{QMC}^{(j)} - \mu)^2\right] = O\left(\frac{(\log N)^{2(d-1)}}{N^2}\right)$$

hence, very small in general, and can be estimated by the standard sample variance estimator

$$\hat{\sigma}_{QMC}^2 = \frac{1}{k-1} \sum_{j=1}^{k} (\hat{\mu}_{QMC}^{(j)} - \hat{\mu}_{QMC})^2.$$

Finally, the algorithm can output an asymptotic confidence interval

$$\hat{I}_\alpha = \left[\hat{\mu}_{QMC} - z_{1-\alpha/2} \frac{\hat{\sigma}_{QMC}}{\sqrt{k}}, \hat{\mu}_{QMC} + z_{1-\alpha/2} \frac{\hat{\sigma}_{QMC}}{\sqrt{k}}\right].$$

6.4 Further readings

There are almost infinitely many books about Monte Carlo methods and several excellent textbooks on Stochastic Simulation. These notes are partially based on [24, 49, 12, 50]; other useful references on Monte Carlo methods are, e.g., [39, 10]. A more advanced presentation of these subject can be found, e.g., in [3]. Regarding sampling and Monte Carlo methods in `R`, a useful reference is [40].